

REMARKS

Claims 1, 2, 4-9, 17-19, and 21-29 are all the claims presently pending in the application. Claims 3, 10-16, and 20 are canceled. Various claims have been amended to more particularly define the invention. Claims 21-29 have been added to claim additional features of the invention. Applicants note that no excess claims fee is due, since the fee for five independent claims has already been paid.

It is noted that Applicants specifically state that no amendment to any claim herein should be construed as a disclaimer of any interest in or right to an equivalent of any element or feature of the amended claim.

Support in the specification for the revised wording of claim 1 is found at lines 8-11 of page 15 and lines 2-5 of page 9. Support for new independent claim 23 is found at lines 7-12 of page 15, line 2 of page 19, lines 6-9 of page 20, lines 12-14 of page 24, and line 6 of page 25. Support for new independent claim 25 is found at lines 3-6 of page 3, line 19 of page 3 through line 3 of page 4, and lines 10-17 of page 4. Support for new independent claim 28 is found at line 16 of page 7 through line 3 of page 8.

Claims 1-20 stand rejected under 35 U.S.C. § 101 as directed toward non-statutory subject matter. Claims 2 and 5-9 stand rejected under 35 U.S.C. § 112, second paragraph, as indefinite. Claims 1, 3, 4, 10, 11, 13-15, and 17-20 stand rejected under 35 U.S.C. § 102(e) as anticipated by U.S. Patent No. 7,031,994 to Lao et al., and claims 2, 5-9, 12, and 16 stand rejected under 35 U.S.C. § 103(a) as unpatentable over Lao.

These rejections are respectfully traversed in the following discussion.

I. THE CLAIMED INVENTION

The claimed invention, as described, for example, in independent claim 1, is directed to a method of increasing efficiency in executing a matrix operation using matrix data in a standard format, the standard format comprising one of a column major format and a row major format. Matrix data stored in the standard format, wherein the matrix data comprises data of any of a complete matrix, a complete submatrix, or a part of a matrix or submatrix, is separated into blocks of data, each block having a size p-by-q. The blocks are rearranged in storage, for retrieval for executing the matrix operation, to be contiguous blocks of contiguous data in a nonstandard format that no longer represents the matrix data in standard

format. The nonstandard format permits the matrix data to be moved from storage into a position for performing the matrix operation more quickly than if the matrix data had been moved in the standard format.

New claim 21 is also directed to the real-world problem of improving computer efficiency.

In another aspect of the present invention described by the new independent claims 26 and 29, the present invention permits a hardware/instruction set deficiency to be overcome by using software instructions only. The software instructions would actually seem to be performing two errors relative to data for the intended operation, but the two errors together are designed to overcome the computer deficiency without having to redesign the machine and by using conventional compilers. In the exemplary embodiment, the deficiency involves the interface with the FPU, the co-processing unit that actually executes the matrix operation described, although the concept is clearly more general.

The present invention is actually only one of several ways used by the inventors to improve efficiency of matrix processing on the assignee's BlueGeneL computer. As explained beginning at line 19 on page 3, the present invention involves a method that improves efficiency and/or overcomes a hardware/instruction set deficiency without going through the expense of a re-design of computer hardware or instructions.

Conventional wisdom would have corrected the deficiency by redesigning the chip in the computer.

The claimed invention, on the other hand, provides a software solution that can be accomplished using conventional compilers and conventional assemblers. Applicants submit that this method is clearly non-obvious, since it involves intentionally executing two errors relative to the intended matrix operation on standard matrix data. That is, the present invention teaches a preliminary processing that converts the matrix data into non-standard format and a new processing added to the matrix operation so that the overall result will be correct for the matrix operation.

However, it is noted that the feature described in independent claim 1 also has utility separate from overcoming a design deficiency, since the standard matrix format using row/column major format is typically a disadvantage for at least one of the three matrices involved in, for example, a matrix multiplication process.

II. THE 35 USC §101 REJECTIONS

Claims 1-20 stand rejected under 35 U.S.C. §101 as allegedly directed to non-statutory subject matter. Applicants respectfully disagree, for the following reasons.

It is noted that a number of the claims have been canceled in this amendment, not because the non-statutory rejection is considered correct, but because Applicants prefer to diversify the definition of the present invention into various aspects not well covered by the original claim set. The following comments are directed only to those claims remaining that are subject to the non-statutory rejection.

First, relative to claims 1-9, the Examiner alleges these claims “... *are directed to a computer method for performing merely numerical computation and manipulation.*” Applicants submit that these claims, even in their original wording, are actually directed toward increasing efficiency of processing linear matrix processing on a computer and/or overcoming a design deficiency at the FPU interface, using only a software remedy. This software remedy includes a preliminary re-arrangement of the data of at least one matrix so that its data is no longer in the standard matrix format using column/row major format.

As explained at line 19 on page 3 through line 4 on page 4, Applicants observed during development of the BlueGeneL computer that this computer architectures failed to have an operation specifically addressed to some stages of matrix processing and that one novel method of overcoming such deficiencies in a computer would be to store the matrix data in a non-standard format (the “register block data format”, see line 18 on page 7) that actually changes order of the information content of the matrix (see “pseudo-matrix” nomenclature in line 20 on page 7) and constitutes a first “error” in the matrix processing. As explained at line 20 on page 7 through line 3 on page 8, a correcting “error” is then intentionally executed that places the information into the computer architecture and instruction set in a manner that “corrects” the first error so that the matrix processing efficiency is improved in a manner that a standard compiler can execute and without having to redesign the computer architecture, components, and/or instruction set.

In the exemplary embodiment discussed in the present invention, the data of the matrix is intentionally placed in “improper” locations as a preliminary processing, thereby changing the order of the information content of the matrix data if this data were to be used as stored. The second “error” in the example of the present invention is defined in claim 2, wherein matrix data from storage is loaded into FPUs in a non-standard manner that corrects

for the dislocation of matrix data, thereby actually increasing efficiency in the matrix processing, while, in the exemplary embodiment, also overcoming an underlying deficiency in the machine architecture and/or instruction set relative to the matrix processing.

Applicants submit, therefore, that this rejection directed to non-statutory subject matter clearly is based upon a misunderstanding of the significance of the claimed invention, since this invention clearly has “real-world” application, by increasing efficiency of processing on a computer, and in the exemplary embodiment, being a part of a mechanism that overcomes one or more inherent deficiencies relative to the matrix processing in that machine.

Thus, Applicants submit that, contrary to the characterization in the rejection, the claimed invention is actually directed toward a method to either increase efficiency of matrix processing in general and/or increase efficiency of matrix processing in a machine that has one or more deficiencies relative to matrix processing. The present invention achieves this increase in efficiency by adding preliminary data processing steps to the data for at least one of the matrices involved in the matrix operation. That is, rather than changing the machine architecture, modifying/ redesigning the machine components, or developing a new machine instruction, the claimed invention describes a method of making corrections for the hardware deficiencies of the machine by converting the matrix data into a format that is different relative to the matrix data stored in standard format.

A correcting “error”, which in the exemplary embodiment, comprises rewriting the matrix operation code to accommodate the nonstandard format by, for example, loading the pseudo matrix data into the FPU register set using optimal loading instructions that can be executed by the processor. Together, these two errors will place the matrix data into the correct location for processing in the FPU, while allowing the data to have been moved faster than if the data had been moved in standard column/row major format for that matrix data. Moreover, the predetermined non standard loading, also a software processing, permits the data to be loaded optimally in the FPU data registers, thereby overcoming a non-optimal hardware design at the FPU interface, using software instructions rather than changing the machine hardware or instruction set.

Therefore, Applicants submit that these claims 1-9, when properly understood, clearly satisfy the “useful, tangible, and concrete” standard required to allow computer-related processes to be statutory subject matter.

Relative to claims 17-19, Applicants submit that these claims are clearly worded to be

directed toward “A signal-bearing medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus” and, therefore, not at all directed to a “... signal carrier which is non-statutory subject matter”, as characterized by the rejection. Rather, these claims are directed toward Beauregard claims and are clearly patentable (see *In re Beauregard*, 53 F.3d 1583 (Fed. Cir., 1995)), as well as US Patent 5,710,578 to Beauregard et al.

In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw these rejections for claims 1-9 and 17-19.

Relative to the new claims, the present invention is generally directed to either improving efficiency of executing a process on a computer or overcoming a hardware/instruction set deficiency of a specific machine configuration by using software instructions that can be handled by conventional compilers. Both these aspects are “real world results” and, therefore, these new claims are clearly directed toward statutory subject matter.

III. THE REJECTION UNDER 35 U.S.C. 112, SECOND PARAGRAPH

The Examiner alleges that claims 2 and 5-9 are indefinite because claim 2 recites “... a format of data ... comprises variations of an optimal floating point loading instruction.” Applicants respectfully disagree, since claim language is required to be interpreted in view of the specification as understood by one having ordinary skill in the art. However, in an effort to expedite prosecution, claim 2 has been amended to address the Examiner’s concern. Therefore, the Examiner is requested to reconsider and withdraw this rejection.

IV. THE PRIOR ART REJECTION

The Examiner alleges that Lao teaches the claimed invention. Applicant submits, however, that there are elements of the claimed invention which are neither taught nor suggested by Lao.

Lao discloses a method to transpose a matrix, as described in the Abstract. Transposition of matrix data is a common matrix procedure that is characterized as preserving the standard format of the matrix. In contrast, the register data format of the present invention places the matrix data into nonstandard format wherein the matrix data is no longer in row major or column major format. The inventors have recognized that storing

matrix data in this nonstandard format can speed up the movement of matrix data to the FPU , as well as potentially being a part of a mechanism that overcomes a deficiency in a specific computer architecture/instruction set relative to matrix processing.

Although there are many permutations of the matrix data that are possible, the present invention teaches that at least one of these many permutations will allow the matrix data for at least one of the matrices involved in a matrix multiply processing to be rearranged in memory so that the matrix data can for that matrix can be retrieved for processing by the FPU as contiguous data for more efficient transport and processing, albeit, in non standard matrix format. This concept of the present invention arises because the standard column/row major format for matrix data will be a disadvantage for retrieving data of at least one of the three matrices involved in matrix multiplication. Additionally, as noted, the present invention addresses a hardware design deficiency at the FPU interface.

That is, as a non-limiting example of a matrix operation, the discussion in the specification addresses specifically matrix multiplication. However, the concepts of the present invention are equally applicable to many other matrix operations and could be extended to more general processing, particularly relative to the aspect of the present invention of overcoming one or more design or interface deficiencies by using software instructions that can be executed with a standard compiler.

Therefore, in contrast to the routine matrix transposition processing in Lao, in one exemplary embodiment, the present invention is directed to the entirely different problem of overcoming a shortcoming in a chip of a computer wherein either the computer architecture or the computer instruction set is deficient, or at least inefficient, relative to the intended matrix processing. The solution proffered by the present invention involves using software that can be implemented by conventional compilers, rather than re-designing the computer chip and/or instruction set.

As defined by the independent claims, the solution involves rearranging the matrix data into a specific nonstandard format such that the matrix data is no longer stored in the standard row major/column major format.

In the exemplary embodiment described in the specification, this rearranging of the matrix data constitutes a first error relative to “normal” matrix processing. In a subsequent step, a second error relative to the intended matrix processing is executed, which in the exemplary embodiment, involves loading the data into the FPU in a cross-loading pattern.

Together, the two errors of the preliminary processing combine in a manner that

corrects the shortcoming of the computer chip.

More specifically, during development of the BlueGeneL computer, the present inventors recognized that the standard memory layout of matrices (e.g., column major or row major order) could to be changed to a layout (e.g., Register Block Data Format, RBDF) prior to the intended processing by Dense Linear Algebra (DLA) routines. These subsequent DLA routines have a common property in that they reuse their data repeatedly while it resides in L1 cache and/or L0 cache. The L0 cache is intended, in the present invention, as referring to the Floating Point Register File (FPRF) of a Floating Point Unit (FPU).

Recently, computer architectures have added special fast floating point units as attachments to their basic processing units. So, to be able to use these units effectively, the present inventors recognized that it is no longer sufficient just to bring data into the L1 cache, since the data should enter L1 and L0 in an optimal seamless manner relative to the new FPUs. The novel RBDF of the present invention has this optimal seamless property whereas conventional column major order or row major order does not.

Hence, as described exemplarily in independent claim 1, the present invention teaches that it pays to reorder standard matrix data layouts, such as CM order, into RBDF before invoking the subsequent DLA routines, as a preliminary processing. Of course, the subsequent DLA routines “know” that their inputs are in RBDF and not standard CM order, as described in dependent claims 2 and 23.

Therefore, a rationale for the present invention now becomes apparent in view of this explanation. Any gain that occurs from using RBDF, instead of CM order, gets multiplied by the repetitive factor inherent in a given subsequent DLA routine that will be using the novel RBDF format of the present invention. Hence, the one-time cost of initially converting the standard CM order into RBDF is quickly paid for. Overall, there is a substantial performance gain in using RBDF, so that the conversion described in the independent claims increases efficiency of the intended matrix operation, while also overcoming the deficiency inherent in newer FPUs by using a software remedy.

Thus, the present invention is not related to the transposition technique described in Lao, since the present invention stores the matrix data in a non-standard format as a preliminary process to the intended matrix operation. A subsequent second error compensates for the incorrect information content. The matrix transposition of Lao is not a preliminary processing, since this transposition processing in Lao is actually the desired matrix operation itself.

Stated slightly differently, the present invention can be viewed as providing a method and structure for high performance processing of linear algebra routines using register block data format (RBDF) routines, wherein conversion into RBDF is a preliminary processing. The register block data format can then be subsequently used by many Dense Linear Algebra (DLA) algorithms. In contrast, Lao, at most, obtains instances of special cases of register block format but subsequently destroys these instances by returning its final output in standard column major (CM) order, and these instances are, therefore, not preliminary processing.

Lao addresses the process of transposing a matrix as the desired matrix processing. In this process, the matrix data starts in standard CM order and output is always returned to standard CM order. In contrast, one aspect of the present invention involves changing the standard CM order input of a matrix A into a new data format RBDF (register block data format).

In paragraph 7 on page 4 of the Office Action, the Examiner refers to Figure 6 of Lao. This figure shows a block diagram of a computer performing the transposition process shown in the flowchart of Figure 5. These figures refer to a special case of a rectangular matrix where M , the number of rows, is related to N , the number of columns, by the formula $N=kM$. The Examiner alleges that "..., each said block having a size 2-by-2; ..."

However, Applicants submit that Lao fails to discuss 2-by-2 blocks in Figure 6 and that the Examiner seems to conclude that this 2-by-2 terminology related to column 16, lines 27-59, where lines 49-53 discuss a checkerboard loading pattern, is similarly related to Figure 6. Applicants submit that the discussion in lines 27-59 of column 16 is related to a specific example for transposing a matrix A with $M=6$ rows and $N=2$ columns.

Applicants submit that it is well known in the art that to transpose a 2-by-2 matrix, one exchanges the elements in the (2,1) and (1,2) positions. This fact is how "checkerboarding" enters into Lao. Lao is forced to use checkerboarding for 2-by-2 submatrices if it is desired to transpose a 2-by-2 submatrix of a larger matrix A .

In contrast, the present invention sometimes will perform a 2-by-2 transposition, perhaps even by using standard out-of-place matrix transposition techniques, to produce the RBDF. However, Lao's purpose is that of a matrix in-place transposition by a faster technique. The present invention has an exemplary purpose of producing a new data format RBDF, for subsequent use by repetitive DLA routines, an entirely different purpose from that of the transposition processing in Lao.

Additionally, Applicants submit that the description in lines 27-59 of column 16 relates to one step of an algorithm described in Figures 7 and 8, along with line 62 of column 16 to line 57 of column 18. An important point is that this is an intermediate step of Figures 7 and 8 and that this instantaneous data layout is subsequently destroyed and returned to standard CM order.

Thus, the present invention clearly differs from Lao, wherein the matrix operation is simple matrix transposition, in that the rearranging of matrix data is a preliminary step to the matrix operation that results in the matrix data being stored for retrieval in a non-standard matrix format. The “matrix operation” in the present invention exemplarily involves a DLA processing.

Hence, turning to the clear language of the claims, in Lao there is no teaching of: “...rearranging and placing in a storage, for retrieval for executing said matrix operation, said blocks of data to be contiguous blocks of contiguous data such that said matrix data is represented in a nonstandard format that permits said matrix data to be moved from said storage into a position for performing said matrix operation more quickly than if said matrix data had been moved as stored in said standard format”, as required by independent claim 1. The remaining rejected independent claims have similar wording and/or wording that clearly distinguishes from the standard format processing of Lao, since Lao is directed toward the process of transposing matrix data.

Therefore, Applicant submits that there are elements of the claimed invention that are not taught or suggest by Lao, and the Examiner is respectfully requested to withdraw this rejection.

V. FORMAL MATTERS AND CONCLUSION

The disclosure has been amended to reflect updated information on the seven co-pending applications, including the present invention, that share a common theme of increasing efficiency of computations of linear algebra processing, which increased efficiency has been a significant contribution on the increased efficiency of the BlueGeneL computer.

In view of the foregoing, Applicant submits that claims 1, 2, 4-9, 17-19, and 21-29, all the claims presently pending in the application, are patentably distinct over the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass

Serial No. 10/671,888
Docket No. YOR920030169US1 (YOR.463)

the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary in a telephonic or personal interview.

The Commissioner is hereby authorized to charge any deficiency in fees or to credit any overpayment in fees to Assignee's Deposit Account No. 50-0510.

Respectfully Submitted,



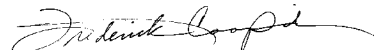
Date: 01/25/07

Frederick E. Cooperrider
Registration No. 36,769

McGinn Intellectual Property Law Group, PLLC
8321 Old Courthouse Road, Suite 200
Vienna, VA 22182-3817
(703) 761-4100
Customer No. 21254

CERTIFICATION OF TRANSMISSION

I certify that I transmitted electronically, via EFS, this Amendment under 37 CFR §1.111 to the USPTO on January 25, 2007.



Frederick E. Cooperrider
Reg. No. 36,769